Introduction

A. Introduction to cloud computing

Over Last few years Cloud computing has emerged as powerful IT technology which provides different IT components and capabilities as services as Pay-as-you-go model. Business organizations are getting benefited by different services provided by cloud. Cloud computing can be defined as "A Technology which provides different computing services like Infrastructure, Platform, Software and Data as a Service." In Cloud computing major participants are cloud service providers and consumers. Both are getting benefited with cloud services as cloud providers own computing resources (e.g. Servers.VMs etc.) and share in among multiple cloud consumers who pay for their usage. On the other hand cloud users use readily available resources and pay on the basis of amount of usage. Thus cloud is a economic model for business organizations.

Cloud has different deployment models public cloud, private cloud, community cloud and hybrid cloud. Public cloud is where cloud resources are shared among multiple independent organizations. Private cloud is a cloud owned by particular organization. Community cloud is owned by group of organizations. Hybrid cloud consists of different types of clouds (public, private, community).

Cloud architecture, involves multiple cloud components like cloud service, platform, infrastructure, storage which are connected and communicate with each other.

Leading organizations in cloud services are Microsoft provides PaaS i.e. Microsoft Azure, Salesforce provides PaaS i.e.force.com, salesforce also offers different applications as CRM, Amazon offers IaaS, Google offers PaaS, GAE and IBM Smartcloud they offer all flavors of cloud services.

B. Challenges in cloud computing- Data security and privacy

Biggest challenge in adoption of cloud technology is data security and privacy. Cloud computing provides cloud storage as a service which stores and manages data for their customers. There is a risk when trusting cloud provider to store important data/files with third party. There are many other issues as in case if user want to change cloud provider, situation when cloud provider close their business and other issues we face while using different services in cloud.

We focus on privacy issue here. As companies start using cloud computing resources for business purposes, business leaders will be wise to find the services they want to use and then review the information security and privacy policies and update them accordingly to address these new risks of privacy. In addition to usage policies for employee interaction on public sites, companies must look for new ways to protect data on resources that are not under their direct control. This includes securing data as it is transmitted to and stored in the cloud as well as granting the appropriate access rights regarding who can view the data.

Privacy is the level of confidentiality provided to user in a system. Privacy not only guarantees the fundamental confidentiality of company data, but also guarantees the data's level of privacy.

Privacy can be violated by intentional release of private company information or through a misapplication of network rights. [2] The threat of data privacy is more in the cloud than traditional technology, due to the number of and interactions between risks and challenges. This is because of the architectural or operational characteristics of the cloud environment. For privacy preservation we need to have strong system which will prohibit unauthorized users from gaining access to sensitive data. Increasing frequency of stolen information and identity theft generated awareness about data security and privacy breaches and their impacts on business. Data loss can have devastating impact on a business or individual. Beyond the damage to one's

brand and reputation, a loss could significantly impact employee, partner and customer moral and trust.

Organizations must take a holistic approach to protecting their information across the enterprise in physical, virtual and cloud infrastructures by:

- Understanding where sensitive data exists
- Safeguarding sensitive data
- Securing and continuously monitoring access to the data
- Demonstrating compliance to pass audits

C. **Need of data privacy in cloud:**

Cloud computing offers economic solutions by providing different services like storage, computing resources etc. There is great demand for sharing person specific data on the internet for various purposes. In today's world more and more historically public information about a person which was discrete is getting linked together to reveal persons entire profile. One major risk of cloud storage is loss of control over personal information. This personal information which is electronically available can be misused.

As large amount of data stored over cloud contains financial or personal sensitive information, storing of this information with third party service provider cause a major threat to privacy of information.

In a cloud data location is dynamic and depends upon various factors such as network speed and availability of storage location. In such a scenario traditional data protection mechanisms which are meant to protect the data at known location, fails due to location uncertainty of user's data.

Review of Literature

A. Access Control Models

S. No	Authors	Topic/ Area of Work(Year)	Methodology used and outcome
1.	Lingfeng C., Doan B.	Novel data protection model in healthcare cloud(2011)	**Methodology:** Cloud based Privacy preserving Role Based Access Control Model (CPRBA) to control and trace data over cloud. Also grant authorized access to cloud resources. Active Auditing Scheme (AAS), it monitors operations and trigger alarm on any operation, data or policy violations in the cloud. **Outcome**: CPRBA provides Fine-grained access control i.e. data owner can define and enforce access control policies on their data. Authorization of data in multi domain environment.AAS prevents policy violation.
2.	Shucheng Yu, Cong Wang, Kui Ren, and Wenjing Lou.	Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing (2010)	**Methodology:** This system defines and enforces access policies based on data attributes. Another advantage is, it enables user to delegate computation tasks to untrusted cloud server without disclosure of any data. This is achieve by combining attribute-based encryption (ABE), proxy re-encryption, and lazy re-encryption. **Outcome:** Highly efficient and provably secure under existing security models
3.	Qun Ni, Alberto Trombetta	Privacy-aware Role Based Access Control(2007)	**Methodology:** Introduced models (P-RBAC), extension of RBAC model. It supports to expressing very complex privacy-related policies, considering different features like purposes and obligations. **Outcome:** Strong privacy access control model for cloud environment
4	Sandhu R., Coyne E., Feinstein, Youman C.	Role Based Access Control Models(1996)	**Methodology:** Presented family of RBAC models. **Outcome:** Models provide a common framework of reference for other research works in this area.

B. Anonymity

Sr. No.	Authors	Topic/Area of Work	Methodology used
1.	Hui Wang.	Privacy-Preserving Data Sharing in Cloud Computing	**Methodology:** A technique, Ambiguity, to protect both privacy with low information loss is proposed. Also proposed an improved scheme, PriView which can achieve better information loss than Ambiguity. **Outcome:** proved effectiveness and efficiency of both Ambiguity and PriView schemes.
2.	P. Samarati, L. Sweeney	Protecting Privacy when Disclosing Information: K-Anonymity and it's enforcement through generalization and suppression	**Methodology:** Illustrates of k-anonymity can be provided using generalization and suppression techniques. **Outcomes:** An approach to disclosing entity specific information such that the released table cannot be reliably linked to external tables
3.	Jian Wang, Yah Zhao, Shuo Jiang, Jiajin Le	Providing Privacy Preserving in cloud	**Methodology:** Anonymity Algorithm used to process data before storing it on cloud server. **Outcomes:** Protects user's privacy in cloud

C. Cryptography

Sr. No.	Authors	Topic/Area of Work	Methodology used
1.	I-Hsun Chuang, Syuan-Hao Li, Kuan-Chieh Huang<yau-Hwang Kuo	An Effective Privacy Protection Scheme for cloud Computing(2011)	**Methodology:** EPPS takes user data privacy requirements and according that derives appropriate composition of algorithm and number of data divisions to preserve user's privacy demand. **Outcomes:** EPPS satisfies user's data privacy requirement and maintains system performance as well

2.	Chhanda Ray, Uttam Ganguly	An approach for Data Privacy in Hybrid cloud environment(2011)	**Methodology:** Data Privacy model defined in which sensitive data is stored on private cloud and non sensitive data over public cloud. Authentication monitor: authenticates request to internal sensitive data grant access to internal authorized requests. **Outcomes:** Authentication monitor: denies access to unauthorized user.
3.	Siani Pearson and Macro Casassa Mont	Sticky Policies: An Approach for managing Privacy across Multiple Parties	**Methodology:** Uses cryptographic mechanism to strongly associate policies with data. The data is encrypted and only accessible upon acceptance and satisfaction of data constraints imposed by policies. **Outcome:** Using this protocol fine grained data access can be achieved by refining definition of policy, encryption and key management.
4.	Angin P., Bhargava B., R.Ranchal , N. Singh, L. Ben Othmane, L. Lilien	An Entity-centric Approach for Privacy and Identity Management in Cloud Computing(2010)	**Methodology:** Defined Two Schemes: 1. Active Bundles: Stores user's identities (identity information), privacy policies and virtual machine. VM enforces the policies and uses a set of protection mechanisms for own protection 2. Anonymous Identification : Enables interactions among the user and cloud services **Outcomes:**1. Active Bundle Scheme: protects privacy of users sensitive data throughout their entire lifecycle and also provide user, control over owned data over cloud. User can decide what and when data will be shared. 2. Anonymous Identification: Shares minimum information with cloud service providers as defined in user's policy. Other characteristics of the approach is, it independent of third party since interaction is only between user and SP.
5.	Gopalakrishnan A.	Cloud Computing Identity Management (2009)	**Methodology:** Survey of cloud security concerns and focused on need of robust Identity Management for cloud services. **Outcomes:** Highlighted need of identity-as-a-service (IaaS)

D. Policy Enforcement

Sr.No.	Authors	Topic/Area of Work	Methodology used
1.	Jun Li, Bryan Stephenson	GEODAC:: A Data Assurance Policy Specification and Enforcement Framework for Outsources Services	**Methodology:** Prototype policy enforcement mechanism using state machine based representation of policies (each state represents a data lifecycle stage) and workflows. **Outcome:** Data policies are defied and enforced.

E. DaaS Data and Computation Privacy

S. No	**Authors**	**Topic/ Area of Work(Year)**	**Methodology used and outcome**
1.	Man Lung Yiu, Ira Assent, Chirstian S.	Outsourced Similarity Search on Metric data Assets(2012)	**Methodology:** obfuscation based privacy, object cannot be distinguished from any other possible object that have same obfuscated representation. **Outcome**: Guarantees correctness of search results and data privacy but at the cost of communication.
2.	C. Curino, E. Jones, R. Popa, N. Malviya, E. Wu, S. Madden, H. Balakrishnan, and N. Zeldovich.	Relational Cloud: A Database Service for the Cloud	**Methodology:** Adjustable privacy using different levels of encryption **Outcome:** Powerful privacy guarantees performance overhead
3.	Yanbin Lu and Gene Tsudik	Privacy-Preserving Cloud Database Querying (2011)	**Methodology:** Encryption scheme **Outcome:** Content-level fine-grained access control by issuing users appropriate search tokens and decryption keys. The scheme also supports blind retrieval of search tokens and decryption keys in the sense that neither data owner nor cloud server learns the query content.

4.	R. Gennaro, C. Gentry, and B. Parno.	Non-Interactive Verifiable Computing: Outsourcing Computation to Untrusted Workers.	**Methodology:** Homomorphic encryption **Outcome:** provides solution to computing on encrypted data but expensive in practice
5.	B. Thompson, S. Haber, W. G. Horne, T. Sander, and D. Yao.	Privacy-Preserving Computation and Verification of Aggregate Queries on Outsourced Databases	**Methodology:** cryptography based solution **Outcome:** provide privacy ware computation but significant query processing and bandwidth consumption
6.	G. Ghinita, P. Kalnis, A. Khoshgozaran , C. Shahabi, and K. L. Tan. .	Private Queries in Location Based Services: Anonymizers are not Necessary.(2008)	**Methodology:** Private information retrieval techniques. **Outcome:** Hide user's query i.e. data item searched for, but not the data queried
7.	K. LeFevre, D. J. DeWitt, and R. Ramakrishnan .	Mondrian Multidimensional K-Anonymity (2006)	**Methodology:** Anonymization technique **Outcome:** secures data using generalization but affect query accuracy
8.	Agrawal, P. J. Haas, and J. Kiernan.	Watermarking Relational Data: Framework, Algorithms and Analysis (2003)	**Methodology:** Digital Watermarking **Outcome:** It establishes owner's identity on data this helps to prove ownership but doesn't prevent illegal copying of data
9.	H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar.	On the Privacy Preserving Properties of Random Data Perturbation Techniques. (2003)	**Methodology:** Perturbation Techniques **Outcome:** Illustrates some of the challenges that these techniques face in preserving the data privacy

Existing literature reviewed as listed in above tables shows contribution to solve privacy issue. Lingfeng C., Doan B.[1] CPBRBA enforces role based privacy aware access control but for some roles that may lead to complete information loss, as there is no provision to display an partial information. They just provide solution to restrict or allow access to sensitive information.

Shucheng Yu and et all proposed scheme [2] provides privacy using cryptographic primitives. This provides secure dynamic resource sharing but has overhead of data encryption and key distribution.

Jui Li, et all presented a framework [14]. This policy framework enables customer easy way to express policy requirements and also enforce those policies using policy enforcement engine. But this provide solution to policy enforcement during data life cycle stages as policies that define requirement during migration, retention etc. But not access level data privacy is not discussed.

Hui Wang. Propose an approach to privacy using concept of ambiguity [8] which is very strong technique to provide data privacy but information loss is an issue.

I-Hsun, et al proposed scheme for user demand privacy and at the same time maintaining system performance. This solution also uses encryption mechanism which will hamper performance.

Chhanda Ray and Ganguly Uttam propose model for data privacy. [12] Only control access to sensitive data using MAC does not guarantee privacy of data.

Siani Pearson and Macro Casassa Mont explained concept of Sticky policies for privacy management. [13] they have suggested a solution to enforce data security using trusted authority. Dependency of third party trusted authority.

In the literature, various approaches for securing DaaS have been studied. Encryption [20,22] provides privacy ware computation but significant query processing and bandwidth consumption. Anonymization[24] secure data with generalization and other approaches but aggregate results may violate data privacy and it also affect query accuracy. Private information retrieval [23] hides user's query requested i.e. data item searched for, but not the

data queried. Watermarking [25] proves the data ownership but does not prevent unauthorized copying of data.

Definition of Problem

Privacy is main hurdle for storing sensitive data on cloud. As financial, healthcare, scientific data are sensitive, requires to be protected. In case such data has to be stored on cloud, it would require privacy protection from unauthorized use. This makes the privacy and authorization on cloud a very important field. Lot of work has already been done to make the cloud more secure by the use of authorization and authentication.

Literature Review shows that there is a need of privacy protection model to safeguard the sensitive data when stored with third party service provider. From literature, proposed to "*Design of a strong privacy model which will also provide considerable improvement in computation cost, time and information accuracy."*

Objectives and issues

Privacy protection techniques applied to data, affect the computation accuracy of results, processing time and also computation cost.

1. To design privacy protection scheme, in order to provide cloud user the confidence to store and retrieve sensitive data on cloud storage.

2. To improve privacy with less compromise in accuracy of results and computation cost.

Scope of Study

This solution could be used in applications where sensitivity of data is vital, such as healthcare, finance, CRM software solutions etc. These applications demand for privacy as sensitive data are maintained by cloud service provider. Our privacy model will provide strong privacy requirements of these applications.

Methodology, tools and techniques:

1. Data collection and experimentation of techniques to find results and compare results for different parameters, performance, and time required for query processing and accuracy of result.
2. Above experimentation on cloud architecture to measure communication overhead.
3. Design a model to protect privacy with minimum effect on computation cost and information accuracy.
4. Experimentation of design proposed in previous stage
5. Testing proposed model for expected parameters as mentioned above.
6. Thesis writing and submission.

Tools and Techniques:

Data collection and analysis tools**,** data hiding techniques & tools: encryption, anonymization, obfuscation, simulation tools.

Plan of Research work (time schedule)

1. **Task:** Data collection and experimentation of techniques to find results and compare results for different parameters, performance, and time required for query processing and accuracy of result.

   Duration: 6 months

2. **Task:** Above experimentation on cloud architecture to measure communication overhead.

   Duration: 6 months

3. **Task**: Design a model to protect privacy with minimum effect on query processing and accuracy.

   Duration: 6 months

4. Experimentation of design proposed in previous stage.

   Duration: 6 months

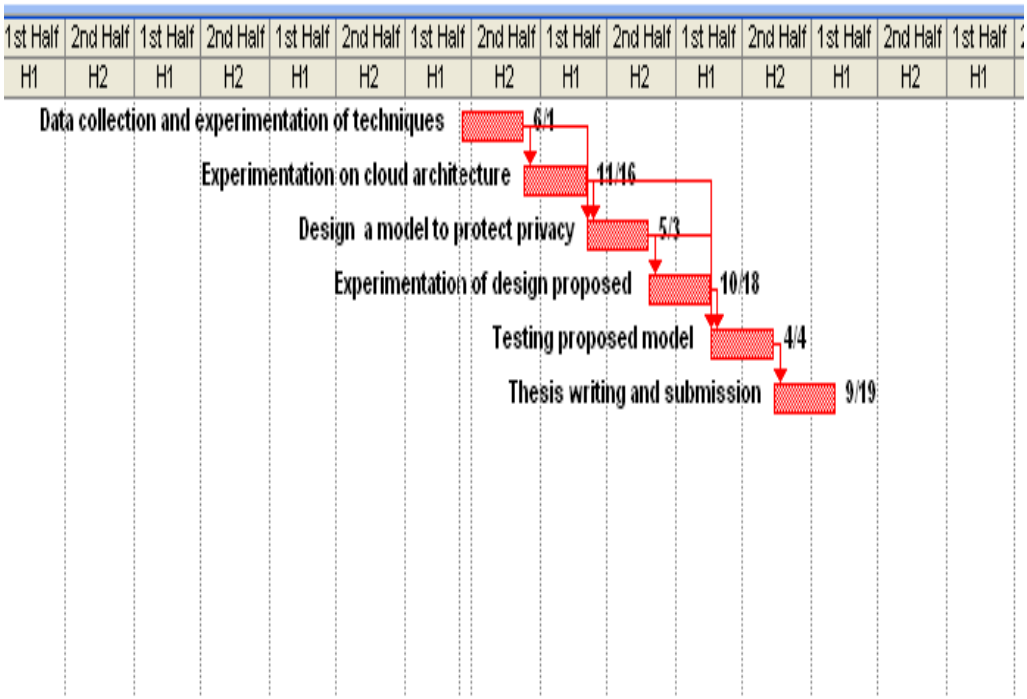
5. **Task:** Testing proposed model for expected parameters as mentioned above.

   Duration: 6 months

6. **Task:** Thesis writing and submission.

   Duration: 6 months

Time Line Chart of Research Plan:



References:

1. Lingfeng C., Doan B., “Novel data protection model in healthcare cloud”, 2011 IEEE International Conference on High Performance Computing and Communications
2. Y. Shucheng, W. Cong, R. Kui, and L. Wenjing, "Achieving Secure,Scalable, and Fine-grained Data Access Control in Cloud Computing" in INFOCOM, 2010 Proceedings IEEE, pp. 1-9, 2010
3. N. Qun, E. Bertino, J. Lobo, and S. B. Calo, "Privacy-AwareRole-Based Access Control" Security & Privacy, IEEE, vol. 7, pp. 35-43,2009.
4. P. Angin, B. Bhargava, R. Ranchal, N. Singh, M. Linderman, L. Ben Othmane, and L. Lilien, "An Entity-Centric Approach for Privacy and Identity Management in Cloud Computing" in Reliable Distributed Systems, 2010 29th IEEE Symposium, pp. 177-183, 2010.
5. W. Cong, W. Qian, R. Kui, and L. Wenjing, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing" in INFOCOM, 2010 Proceedings IEEE, pp. 1-9, 2010
6. R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models" Computer, vol. 29, pp. 38-47, 1996.
7. P. Angin, B. Bhargava, R. Ranchal, N. Singh, M. Linderman, L. BenOthmane, and L. Lilien, "An Entity-Centric Approach for Privacy and Identity Management in Cloud Computing" in Reliable Distributed Systems, 2010 29th IEEE Symposium, pp. 177-183, 2010.
8. Hui Wang. Privacy-preserving data sharing in cloud computing. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY, May 2010

9. Samarati P, Sweeney L. Protecting privacy when disclosing information: *k*-anonymity and its enforcement through gen-eralization and suppression. Technical Report, SRI International, 1998.
10. I-Hsun Chuang, Et al, An effective privacy protection scheme for cloud computing, February 13,2011 , ISBN 978-89-5519-155-4
11. Jian Wang, Yan Zhao, Shuo Jiang, Jiajin Le, "Providing Privacy Preserving in cloud computing", Proceedings of International Conference on Test and Measurement, 2009, pp 213 – 216
12. Chhanda Ray,Ganguly Uttam, "An Approach for Data Privacy in Hybrid Cloud Environment", International Conference on Computer & Communication Technology (ICCCT)-2011
13. Siani Pearson and Macro Casassa Mont," Sticky Policies: An Approach for managing Privacy across Multiple Parties" Hp Labs Bristol, IEEE Computer Society, 2011.
14. Jun Li, Bryan Stephenson, "GEODAC:: A Data Assurance Policy Specification and Enforcement Framework for Outsources Services", IEEE transaction on services computing, vol 4, No. 4, OCT-DEC 2011.
15. Robert Gellman, Privacy in Cloud: Risk to Privacy and Confidentiality from cloud computing, World privacy Forum, February 23,2009, pp 4-7 http://www.worldprivacyforum.org/pdf/WPF_Cloud_Privacy_Report.pdf
16. Dawei Sun, *Et al*, Surveying and Analyzing Security, Privacy and Trust Issues in cloud computing environments, ScienceDirect, published by ELSEVIER, 2011.
17. Cloud Security Alliance, Top Threats to Cloud Computing, March 2010. https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf

18. Man Lung Yiu, Ira Assent, Christian S. ”Outsourced Similarity Search on Metric Data Assets”, IEEE TRANSACTION AND DATA ENGINEERING, VOL 24., NO.2, FEBRUARY 2012.
19. C. Curino, E. Jones, R. Popa, N. Malviya, E. Wu, S. Madden, H. Balakrishnan, and N. Zeldovich. Relational Cloud: A Database Service for the Cloud. In CIDR, pages 235--240, 2011
20. Yanbin Lu and Gene Tsudik *“ Privacy-Preserving Cloud Database Querying*” Journal of Internet Services and Information Security (JISIS), 1(4): 5-25, November 2011
21. R. Gennaro, C. Gentry, and B. Parno. Non-Interactive Verifiable Computing: Outsourcing Computation to Untrusted Workers. STOC, 2010.
22. B. Thompson, S. Haber, W. G. Horne, T. Sander, and D. Yao.Privacy-Preserving Computation and Verification of Aggregate Queries on Outsourced Databases. HPL-2009-119, 2009.
23. G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K. L. Tan. Private Queries in Location Based Services: Anonymizers are not Necessary. In SIGMOD, pages 121–132, 2008
24. K. LeFevre, D. J. DeWitt, and R. Ramakrishnan. Mondrian Multidimensional K-Anonymity. In ICDE, page 25, 2006.
25. Agrawal, P. J. Haas, and J. Kiernan. Watermarking Relational Data: Framework, Algorithms and Analysis. VLDB J., 12(2):157–169, 2003
26. H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar. On the Privacy Preserving Properties of Random Data Perturbation Techniques. In ICDM, pages 99–106, 2003.